

AMENDMENTS TO THE CLAIMS:

This listing of claims will replace all prior versions, and listings, of claims in the application. Please amend claims 1 and 16 and add new claims 21-28 as follows:

LISTING OF CLAIMS:

1. (Currently Amended) A method for extracting information from a natural language text corpus based on a natural language query, comprising the steps of:
analyzing said natural language text corpus with respect to surface structure of word tokens and surface syntactic roles of constituents;
A
indexing and storing the analyzed natural language text corpus;
analyzing a natural language query with respect to surface structure of word tokens and surface syntactic roles of constituents;
creating one or more a number of surface variants of the analyzed natural language query, ~~said one or more by replacing word tokens of said natural language query, and for at least one surface variant by rearranging word tokens of said natural language query, in such a way that said number of~~ surface variants being are equivalent to said natural language query with respect to lexical meaning of word tokens and surface syntactic roles of constituents;
comparing said one or more number of surface variants and said analyzed natural language query with the indexed and stored analyzed natural language text corpus; and

extracting from said indexed and stored analyzed natural language text corpus, each portion of text comprising a string of word tokens that matches any one of said surface variants or said analyzed natural language query.

2. (Original) The method according to claim 1, wherein, in the step of creating said surface syntactic roles of constituents are head and modifier roles, and grammatical relations.

3. (Original) The method according to claim 1, wherein, in the step of extracting, a string of word tokens in said indexed and stored analyzed natural language text corpus matches one of said surface variants or said analyzed natural language query if it comprises the head words of phrases bearing the grammatical relations of subject, object, and lexical main verb in said one of said surface variants or said analyzed natural language query in the same linear order as in said one of said surface variants or said analyzed natural language query.

4. (Original) The method according to claim 1, wherein, in the step of analyzing a natural language query, said natural language query is analyzed in the same manner as said natural language text corpus is analyzed in the step of analyzing said natural language text corpus.

5. (Original) The method according to claim 1, wherein the step of analyzing a natural language text corpus comprises the steps of:

determining a morpho-syntactic description for each word token of said natural language text corpus;

locating phrases in said natural language text corpus;

determining a phrase type for each of said phrases;

locating clauses in said natural language text corpus,

and wherein the step of analyzing a natural language query comprises the steps of:

determining a morpho-syntactic description for each word token of said natural language query; and

locating phrases in said natural language query;

determining a phrase type for each of said phrases; and

locating clauses in said natural language query.

6. (Original) The method according to claim 5, wherein the step of indexing and storing comprises the steps of:

providing, for each word token of said natural language text corpus with, a unique word token location identifier;

storing information regarding the location of each word token of said natural language text corpus, based on said unique word token location identifiers;

storing, for each phrase type, information regarding the location of each phrase of this type in said natural language text corpus, based on said unique word token location identifiers; and

storing information regarding the location of each clause in said natural language text corpus, based on said unique word token location identifiers.

7. (Original) The method according to claim 6, wherein each word token is associated with a word type, and wherein the step of storing information regarding the locations of each word token comprises the steps of:

storing each word type of said natural language text corpus; and
storing, for each word token, its unique word token location identifier logically linked to the stored associated word type.

8. (Original) The method according to claim 7, wherein the step of storing information regarding the locations of phrases comprises the steps of:

providing, for each phrase of said natural language text corpus, a unique phrase location identifier identifying the word tokens spanned by the phrase;
storing each phrase type of said natural language text corpus; and
storing, for each phrase, its unique phrase location identifier logically linked to the stored associated phrase type.

9. (Original) The method according to claim 8, wherein the step of storing information regarding the locations of clauses comprises the steps of:
providing, for each clause of said natural language text corpus, a unique clause location identifier identifying the word tokens and phrases spanned by the clause;
storing, for each clause, its unique clause location identifier.

10. (Original) The method according to claim 9, further comprising the steps of:
locating sentences in said natural language text corpus; and
providing, for each sentence of said natural language text corpus, a unique sentence location identifier identifying the word tokens, phrases and clauses spanned by the sentence;
storing, for each sentence, its unique sentence location identifier.

11. (Original) The method according to claim 10, further comprising the steps of:
locating paragraphs in said natural language text corpus;
providing, for each paragraph of said natural language text corpus, a unique paragraph location identifier identifying the word tokens, phrases, clauses and sentences spanned by the paragraph;
storing, for each paragraph, its unique paragraph location identifier.

12. (Original) The method according to claim 11, further comprising the steps of:
locating documents in said natural language text corpus;
providing, for each document of said natural language text corpus, a unique
document location identifier identifying the word tokens, phrases, clauses, sentences and
paragraphs spanned by the document;
storing, for each document, its unique document location identifier.

13. (Original) The method according to claim 1, wherein, in the step of
extracting, a portion of text that is extracted is either the matching string of word tokens, a
clause comprising the matching string of word tokens, a sentence comprising the matching
string of word tokens, a paragraph comprising the matching string of word tokens, or a
document comprising the matching string of word tokens.

14. (Original) The method according to claim 1, further comprising the step of:
organizing the extracted information according to degree of correspondence with the
query with respect to lexical meaning of word tokens and surface syntactic roles of
constituents, such that a constituent in a portion of text having the same lemma as the
equivalent constituent of the query is considered to have a higher degree of correspondence
than a constituent in a portion of text being a synonym to the equivalent constituent of the
query.

15. (Original) The method according to claim 1, further comprising the step of: organizing the extracted information such that said portions of text are grouped according to sameness of grammatical subject, grammatical object, and lexical main verb.

16. (Currently Amended) A system for extracting information from a natural language text corpus based on a natural language query, comprising:
a text analysis unit for analyzing a natural language text corpus and a natural language query with respect to surface structure of word tokens and surface syntactic roles of constituents;

A
storage means operatively connected to said text analysis unit, for storing the analyzed natural language text corpus;

an indexer, operatively connected to said storage means, for indexing the analyzed natural language text corpus;

an index, operatively connected to said indexer, for storing said indexed analyzed natural language text corpus;

a query manager, operatively connected to said text analysis unit, comprising means for creating surface variants of said natural language query[[,]] by replacing word tokens and rearranging word tokens of said natural language query in such a way that said surface variants being are equivalent to said natural language query with respect to lexical meaning of word tokens and surface syntactic roles of constituents, and means for comparing said

surface variants and said analyzed natural language query with the indexed analyzed natural language text corpus in said index; and

a result manager operatively connected to said index, for extracting, from said indexed and stored analyzed natural language text corpus, each portion of text comprising a string of word tokens that matches any one of said surface variants or said analyzed natural language query.

17. (Original) The system according to claim 16, wherein a string of word tokens in said indexed and stored analyzed natural language text corpus matches one of said surface variants or said analyzed natural language query if it comprises the head words of phrases bearing the grammatical relations of subject, object, and lexical main verb in said one of said surface variants or said analyzed natural language query in the same linear order as in said one at said surface variants or said analyzed natural language query.

18. (Original) The system according to claim 16, wherein said index comprises multiple indexes based on a hierarchy of text units that are related by inclusion.

19. (Original) A computer readable medium having computer-executable instructions for a general-purpose computer to perform the steps recited in claim 1.

20. (Original) A computer program comprising computer-executable instructions for performing the steps recited in claim 1.

21. (New) A method of storing a natural language text corpus in a database, comprising the steps of:

identifying word tokens of said natural language text corpus;

determining locations in the natural language text of the identified word tokens;

determining word types associated with the identified word tokens;

storing the determined word types in said database, wherein the number of stored word types is less than the number of identified word tokens;

storing word token location identifiers identifying the determined locations in the natural language text corpus of the identified word tokens; and

linking the stored word token location identifiers to the stored word types, such that, for a given identified word token, the stored word token location identifier identifying the location of the identified word token is logically linked to the stored word type associated with the identified word token.

22. (New) The method according to claim 21, further comprising the steps of:

determining morpho-syntactic descriptions for the identified word tokens;

storing the morpho-syntactic descriptions for the identified word tokens; and

linking the stored morpho-syntactic descriptions to the stored word token location identifiers, such that, for a given identified word token, the stored morpho-syntactic description for the identified word token is logically linked to the stored word token location identifier identifying the location of the identified word token.

23. (New) The method according to claim 22, wherein the morpho-syntactic description of a word token comprises a part-of-speech and an inflectional form of the word token.

24. (New) The method according to claim 21, further comprising the steps of:
identifying phrases of said natural language text corpus;
determining word tokens comprised in the identified phrases; and
storing phrase location identifiers identifying the stored word token location
identifiers of the word tokens comprised in the identified phrases, such that, for a given
identified phrase, the stored phrase location identifier of the identified phrase identifies the
stored word token location identifiers identifying the location of the word tokens comprised
in the identified phrase.

25. (New) The method according to claim 24, further comprising the steps of:
determining phrase types of the identified phrases;
storing the determined phrase types; and
linking the stored phrase types to the stored phrase location identifiers, such that, for
a given identified phrase, the phrase type for the identified phrase is logically linked to the
stored phrase location identifier identifying the stored word token location identifiers
identifying the location of the word tokens comprised in the identified phrase.

26. (New) A system for storing a natural language text corpus, comprising:
a text analysis unit for identifying word tokens of said natural language text corpus,
determining locations in the natural language text of the identified word tokens, and
determining word types associated with the identified word tokens;
a database for storing the determined word types, wherein the number of stored word
types is less than the number of identified word tokens, storing word token location
identifiers identifying the location in the natural language text corpus of a respective
identified word token, and linking the stored word token location identifiers to the stored
word types, such that, for a given identified word token, the stored word token location
identifier identifying the location of the identified word token is logically linked to the
stored word type which is associated with the identified word token.

27. (New) The system according to claim 26, wherein the text analysis unit is
further adapted to determine morpho-syntactic descriptions for the identified word tokens,
and the database further stores the morpho-syntactic descriptions for the identified word
tokens, and links the morpho-syntactic descriptions to the stored word type location
identifiers, such that, for a given identified word token, the morpho-syntactic description
for the identified word token is logically linked to the stored word token location identifier
identifying the location of the identified word token.

28. (New) The system according to claim 27, wherein the morpho-syntactic
description for the word token comprises a part-of-speech and an inflectional form of the
word token.
